DOCUMENT RESUME

ED 027 097 RC 001 862

By-Palomares, Uvaldo Hill

A Critical Analysis of the Research on the Intellectual Evaluation of Mexican-American Children.

Pub Date May 65

Note-51p.; Term paper presented to School of Education, University of Southern California

EDRS Price MF-$0.25 HC-$2.65

Descriptors-Anglo Americans, *Comparative Statistics, English (Second Language), *Intelligence Differences, Intelligence Factors, *Intelligence Tests, Measurement Techniques, *Mexican Americans, Research Criteria, *Research Reviews (Publications), Spanish Americans

The validity is questioned of testing the intelligence of Mexican Americans via intelligence tests constructed for Anglo Americans. An overview is included of the literature concerned with the comparability of intelligence scores. Critiques are presented of 9 current studies utilizing intelligence test scores of Mexican Americans and Anglo Americans as a basis for comparative analysis. It is suggested that amount of acculturation is closely related to other variables involved in intelligence and should be investigated as a possible contaminating variable in measuring the intelligence of Mexican Americans. (JAM)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

A CRITICAL ANALYSIS OF THE RESEARCH ON THE INTELLECTUAL EVALUATION OF MEXICAN-AMERICAN CHILDREN

A Critique

Presented to

Dr. Kenneth Hopkins

School of Education

University of Southern California

In Partial Fulfillment

of the

Requirements for the Course

Ed. Ps. 792b

by

Uvaldo Hill Palomares

May, 1965

ED 021962
RC 001 862

TABLE OF CONTENTS

PAGE

I. ORIENTATION TO SUBJECT AREA AND STRUCTURE OF CRITIQUE 1

Introduction 1

Purpose 2

Overview 2

Discussion of the term Mexican-American 2

Background 3

Structure of the Critique 4

II. THE STUDIES CRITIQUED 5

Category 1 5

Language Barrier as an Educational Problem of Spanish-Speaking Children, by Holland, W. R. 6

A Study of the Performance on English and Spanish Editions of the Stanford-Binet Intelligence Test by Spanish-American Children, by Keston, M. J. & Jimenez, C. 11

The Value of the Leiter Scale in Testing Mexican-American Children, by Burns, P. B. 13

Category II 15

Intellectual Functioning of the Mentally Retarded, by Silverstein, A. B.; Shotwell, A. M.; & Fisher, G. M. 16

Intelligence of American Children of Mexican Parentage, by Carlson, H., & Henderson, Norman 19

An Evaluation of the Davis-Eells (Culture-Fair) Test Using Spanish and Anglo-American Children, by Stablin, J. E., Willey, D. S., & Thompson, C. W. 24

W.I.S.C. Patterns of a Selective Sample of Bilingual School Children, by Altus, G. T. 28

Category III..... 31

The Culture-Fair Aspects of Cattell's Test of g: Culture-Free, by Kidd, A. H. 32

Learning Abilities in Mexican-American and Anglo-American Children, by Jensen, A. R. 37

III. SUMMARY AND CONCLUSIONS..... 44

IV. BIBLIOGRAPHY..... 47

ORIENTATION TO SUBJECT AREA

AND STRUCTURE OF CRITIQUE

Introduction

The intellectual measurement of the Mexican-American (M-A) has presented a challenge and a problem to those interested for many years. The unique environmental millieu in which he develops seems loaded with contaminating variables that lead to problems in the assessment of his intelligence. The need for clarification of the intellectual ability of the M-A is all the more apparent when one becomes aware of the large number of M-As living throughout the southwest and in California, in particular. The 1960 Census of Population revealed that there were 1,426,538 persons of Spanish surname living in California at that date. This figure comprises 9.1% of the population of the state. The problem of measuring the intellectual ability of the M-A becomes more apparent when one considers that this population grew much more rapidly between 1950 and 1960, by 88.1%, than did the state's total population, which grew by 48.5% (Calif. State Report, 1964). Although all individuals of Spanish surname are not to be included in the problematic section of the population being discussed here, the figures presented do give some indication of the number of people involved.

There are programs in action now and new programs being developed that call for the evaluation of the M-A population by instruments which may be penalizing them. All programs in the schools call for the selection of students by the use of psychometric devices, as do the mandatory special classes for the mentally retarded, which many times have been found to be overly populated with M-As, because of the inadequate testing devices used (Burnes, 1955). New programs supported by the state and the Federal Government, such as the English as a Second Language Program and the Compensatory Education Program are also beginning to make demands for psychometric devices for the evaluation of the intellectual ability of the M-A.

Purpose

The purpose of this Critique is to explore the Intellectual evaluation of the M-A in the following manner:

1. To first present a concise overview of the problem in terms of the literature in the field.

2. To select and critique noteworthy studies which might bring the reader up-to-date on the present state of the research concerning (1) the state of knowledge in the area, and (2) the methodological problems and procedures involved in assessing the problem in question.

3. To arrive at some summarizing conclusions from the total report and to recommend a specific area for further study.

Overview

The use of intelligence tests in studies with different racial and ethnic groups has generally resulted in lower scores among immigrant and other minority groups than among comparable, native born Anglo-Americans* (Kidd, 1952). This generalization holds true for the selection of the population which is referred to as Mexican-American, Spanish-American, Mexican, individuals of Mexican parentage, etc. Because of the confusion involved in defining the population with which this paper is concerned, no exact definition will be attempted, rather a general discussion of the definitional complexities of the term Mexican-American will be presented.

Discussion of the term Mexican-American

The term Mexican-American has been loosely used to describe any individual of Spanish surname, as was done in the introductory part of this paper. This use of the term is obviously too wide because it includes any individual who comes from any country that has within its borders individuals of Spanish Surname. When we exclude individuals from countries other than Mexico, we are confronted with the problem of whether the individual was born in the United States or in Mexico. The term Mexican-American has its semantic problems as well, because

*The term Anglo-American (A-A) is used throughout this paper to refer to white, American-born children from monolingual, English-speaking families.

individuals living in Mexico could also be considered Mexican-American by virtue of living in Mexico which is in the Americas.

Assuming that we were to limit ourselves to those individuals born in Mexico or the United States, but presently residing in the United States, the definitional dilemma would narrow but not necessarily lessen. The M-A subculture has many intra-group differences which contribute to the difficulty in trying to define the total group or a part of it.

Because of the breadth and diversity in background, design, sophistication of the studies and problems involved in defining the M-A subculture, operational definitions have been supplied by the authors of the studies critiqued in this report and are presented herein.

Background

The seemingly poor aptitude and achievement of the M-A children has been especially conspicuous to public school personnel in parts of California. (Jensen, 1961; Record, 1959). Of special significance is the observation (Jensen, 1961; Burnes, 1955) that in a number of California school districts many of the M-A children are classified as being "slow learners" and/or "mentally retarded" on the basis of their scores received on currently popular, standardized intelligence tests. Many of these children are placed in special classes for the mentally retarded.

Slow learners are not the exception, but a majority of the M-A students (Holland, 1960). Verbal IQ tests have tended to classify the M-A student as being considerably below the average of the predominately A-A normative groups on which the tests were standardized. (Kidd, 1962; Coers, 1935; Keston, 1954). The observation that M-As score as a group below the A-A has, however, been found for both verbal and non-verbal tests. (Hill, 1936; Stablin, 1961). Previous investigations (Bingham, 1930; Daniel, 1932; Koch, 1926; Keston, 1954) show that the IQ of M-A children averages from eleven to twenty-two points below A-A

children. The differences between these groups were somewhat less if non-language tests are used. (Carlson, 1950). It is interesting to note that the Spanish translation of the Stanford-Binet has failed to make any significant differences in the gap between M-A and A-A groups. Both the English and Spanish versions of the Stanford-Binet tend to cluster the average IQ of the M-A around 80.9. (Coers, 1935; Cook, 1955; Hill, 1936).

Structure of the Critique

In reviewing the relevant literature, certain characteristic modes of research began to emerge which will be used as guidelines for this critique. The reader should be aware that the following organization of the literature is an oversimplification and has been attempted only to facilitate the structure of this critique. Further, that the categories used herein are not intended as being mutually exclusive. For the purpose of effecting clear organization in this critique, the studies have been grouped in three categories:

1. The exploration of the intelligence of the M-A via the use of one experimental M-A group and two IQ tests; one IQ test with the ability to measure "capacity," and the other unable to do so.

2. The exploration of the intelligence of the M-A via the use of two groups; a control group composed of A-As, and an experimental group of M-As. The comparison is on the basis of one or more instrument. The direct evaluation is of the intelligence of the M-A group, the testing instrument, or in some cases, both.

3. The exploration of the intelligence of the M-A via the use of four groups with the critical evaluation centered on the value of the instrument or instruments for assessing the intellectual ability of the M-As.

Category I

Study of the M-A IQ via 1 Group and 2 Tests

Three general approaches are used in this category for the study of the intelligence of the M-A. In the first approach, the researcher takes one experimental population and administers to it one test, in most cases it is the Weschler Intelligence Scale for Children (WISC). This is done because of the verbal and non-verbal scales in the WISC which can be treated as separate IQs. In this manner, the researcher can compare the verbal and non-verbal scores and arrive at some conclusions concerning the M-As IQ.

A second variation, of essentially the same method, is done by taking one experimental group and administering to it two versions of the same IQ test. The tests used extensively in this method are the English and Spanish versions of the Stanford-Binet IQ test. There are many Spanish versions presently available. The examiner uses one of these versions, or translates one himself, and administers it to the experimental group along with an English form of the same test. The comparison is then interpreted as revealing of the M-As intelligence.

A third approach, related somewhat to the former methods is done by comparing the findings of two IQ tests that are unrelated. Usually the results of a non-verbal IQ test are compared to the results of a verbal test for the same M-A population. This comparison is interpreted as revealing of the M-A's intelligence.

Three studies will be critiqued in this first section with the intention of presenting a sample of each one of the approaches described above. The samples will appear in the same order as described in the above paragraphs.

Sample of Approach a (comparison of W-Ac IQs on verbal and non-verbal scales of same test).

Holland, William R., "Language Barrier as an Educational Problem of Spanish-Speaking Children," Exceptional Children, 27, pp. 42-50, 1960.

Purpose

This study attempted to define and analyze the social and cultural background of the educational problems of 36 Spanish-speaking children who were recommended for psychological testing in the Tucson Public School System in the spring of 1958.

Method

Sample: The Weschler Intelligence Scale for Children (WISC) was administered bilingually to 36 Spanish-speaking children ranging from grades 1C (remedial class to remedy linguistic handicap of Spanish-speaking children before placement in the regular 1st grade) through grade 5. All were born and had spent all of their lives in Tucson or in other parts of southern Arizona. The study pop. represents a sample of the Spanish-speaking children tested in the Tucson schools but is not a representative sample of total Spanish-speaking pop. of the Tucson Schools.

Author might have included information concerning sex, and age distribution.

Testing Procedure: The WISC instructions, questions and answers were translated into Spanish and mimeographed together with a description of the method of bilingual administration in a manner developed by the author. Bilingual administration consisted of giving the instructions for the WISC subtests in English first. Only when the instructions were not understood or were only partially understood in English were they repeated in Spanish.

This method was also used in presenting the questions of the 5 verbal subtests in the WISC. Correct answers in either language were accepted and given credit. This procedure yielded two types of responses, English R (correct answer in English) and Bilingual R (correct answer in Spanish). The raw score totals of these subtests were computed separately for the two classes of responses.

Any change in the wording or test administration procedures of a test like the WISC would render the generalizations from the standardizing pop. to the present one questionable.

It is not made clear if the rules dictating that after so many errors a subtest should be discontinued. This is an interesting point since the ability in Spanish might allow some of these Ss to try a question in English even after they had missed their allotted number because of their ability to carry through their answers in Spanish

No language distinction was made in scoring the Digit Span subtest or the perf. section, which were administered by the conventional WISC method.

Interpretive Concepts: The system of correction described above yields two verbal scores, an English Verbal IQ and a Bilingual Verbal IQ. The first represents the present level of functioning in English language skills. It indicates most closely the present level of language achievement in the school situation. The second suggests the future potential for verbal skills when the subjects' knowledge of English is approximately equal to that of Spanish. The IQ point difference between these scores points up the value of the Language Barrier. The formula is:

Language Barrier = Bilingual Verbal IQ - English Verbal IQ

The Perf. IQ was believed to represent the approximate intellectual potential under optimal social and cultural conditions. A comparison of the perf. IQ with the Bilingual Verbal IQ serves to point up how near the subjects' present level of functioning in a bilingual environment is to his potential under conditions more favorable to language development.

Results

The WISC results for the Ss are summarized in Table 1. The range of the English Verbal IQ scores was from 45 to 118 points and the mean for the group was 85.2. This yielded an average Language Barrier of 4.6 IQ points per student.

Table II describes the Language Barrier in IQ points. There was a relationship between the amount of Language Barrier and the degree of difficulty encountered in academic achievement. The SD of 4.05 indicates that about 68% of the Ss can be expected to have up to 9 points of Language Barrier.

Were test instructions on the Perf. and Digit Span given all in English with no help in Spanish?

The validity of using an IQ test for the described interpretive concepts is questioned on the following points:

1. The interpretation of the Bilingual Verbal score as indicate of the potential for future verbal skills is questioned on the basis that these children have been found to be deficient in both languages (Keston, 1954) and therefore such a measure would very likely, if at all valid, measure their present state of language ability in Spanish but certainly not their potential;

2. Since in the normal pop. on which the WISC was standardized many times verbal scores are in excess of perf. scores. The use of the perf. score to measure optimal functioning intellectual ability may many times sell a subject short of his "true" potential.

No information was presented which demonstrated the relationship between Language Barrier and academic achievement.

Table 1

WISC Results for 36 Spanish-Speaking Children Tested Bilingually

Subject	Gr. Level	Eng. Verb. IQ	Bil. Verb. IQ	Lang. Barr. in IQ Pts.	Perf. IQ	Diff. Perf. IQ & Eng. Verb IQ	Diff. Perf. IQ & Bil. Verb IQ	Full Scale Bil. IQ
1	1C	52	57	5	50	-2	-7	49
2	1C	46	48	2	53	7	5	45
3	1C	63	81	18	103	40	22	91
4	1C	60	67	7	78	18	11	70
5	1C	75	95	20	83	8	-12	88
6	1C	90	67	7	75	15	8	68
7	1	113	115	2	111	-2	-4	115
8	1	95	97	2	106	11	9	101
9	1	109	110	1	99	-10	-11	105
10	1	82	96	14	94	12	-2	95
11	1	74	77	3	90	16	13	82
12	1	57	67	10	85	28	18	73
13	2	69	74	5	78	9	4	78
14	2	80	84	4	115	35	31	99
15	2	72	74	2	72	0	12	71
16	2	69	72	3	83	14	11	75
17	2	61	69	8	85	24	16	74
18	2	80	90	10	89	9	-1	88
19	3	74	80	6	93	19	13	85
20	3	76	79	3	74	-2	-5	74
21	3	79	82	3	86	7	4	83
22	3	72	77	5	76	4	-1	75
23	3	81	82	1	106	25	24	93
24	3	94	95	1	101	7	6	98
25	4	65	70	5	82	17	12	73
26	4	74	77	3	79	5	2	75
27	4	95	95	0	87	-8	-8	91
28	4	85	87	2	83	-2	-4	84
29	4	89	91	2	79	-10	-12	84
30	4	89	89	0	117	28	28	102
31	5	104	106	2	106	2	0	107
32	5	118	118	0	115	-3	-3	118
33	5	97	99	2	99	2	0	99
34	5	110	111	1	122	12	22	118
35	5	92	96	4	122	30	26	109
36	5	94	95	1	94	0	-1	94

Table I

(Continued)

	E.V. IQ	B.V. IQ	L.B. in IQ pts.	P IQ	D. P. IQ E. V. IQ	D. P. IQ B. V. IQ	Ful S. B. IQ
M	80.6	85.2	4.6	90.8	+10.2	5.6	86.7
SD	16.2	16.7	4.05	17.4	13.6	13.5	16.5

Table II

Language Barrier in IQ points

Points	Students	Language Barrier
0	3	None
1-3	18	Moderate
4-6	7	Serious
7+	8	Very Serious

Table III

WISC Results Related to the Grade Levels of 36 Spanish-Speaking Children

Grade	1C	1	2	3	4	5
Number of Students	6	6	6	6	6	6
Mean English Verbal IQ	59	88	72	79	83	103
Mean Bilingual Verbal IQ	69	94	77	83	85	104
Mean Performance IQ	71	98	87	89	88	110
Mean Language Barrier	10	5.3	5.3	3.2	2	1.7
Median Lang.	7	2.5	4.5	3	2	1.5
Barrier (Range)	2.20	1-14	2-10	1-6	0-5	0-4

Language difficulties diminish steadily with each successive year of schooling, being greatest in 1C children. Most of the first and second graders tested still had moderately high Language Barriers. (See table III.)

There was a mean difference of 10.2 between the English Verbal IQ scores and the Perf. IQ scores favoring the latter. The difference of the two scores (C. R. = 4.9) was significant at the 1% level of confidence.

The Perf. results were on the average 5.6 points higher than the Bilingual Verbal IQ. The diff. (C. R. = 2.8) was significant at the 1% level of confidence.

Conclusions

1. A child has a Language Barrier when his knowledge of Spanish is greater than his knowledge of English.

2. There was an average Language Barrier of 4.6 IQ points per student. In 8 cases the Language Barrier was very serious; in 7 cases it was serious; and in 18 cases it was moderate. Only 3 students were completely free of Language Barrier.

3. Although Language Barrier declines with increased schooling, it is still present among some fifth graders. Many first, second and third graders still have serious and very serious Language Barriers.

4. Although grade 1C reduces Language Barrier considerably, many students beyond this level still have serious language problems.

5. Language Barrier is a very important factor in the lower academic achievement of many Spanish-speaking children in the Tucson Public Schools.

This is an interesting approach toward an objective index of the bilingual, Spanish-speaking child's language handicap in an academic setting. At present there are many English as a Second Language Programs being conducted in California. Their most pressing, single problem is to find an instrument which would identify a child with a linguistic handicap well enough in order that he be placed correctly. To date the decisions have been made mostly on teacher judgement, which many times leads to conflicts with other educational programs being conducted, such as the EMR program. The approach suggested by this study is certainly worth considering in and of itself, or as a guideline for a more valid instrument.

Sample of approach b (comparison of M-A IQ on Spanish and English translation of same test).

Keston, Morton J. and Jimenez, Carmina, "A Study of the Performance on English and Spanish Edition of the Stanford-Binet Intelligence Test by Spanish-American Children," Journal of Genetic Psychology, 85, pp. 263-269, 1554.

Purpose

The purpose of the study was to compare the intelligence core of the subjects in the one language with their intelligence test score in the other language and to discuss some reasonable explanations for any difference found. A specific purpose of the study was to determine whether the bilingual children of Albuquerque, New Mexico, should be given the Stanford-Binet Intelligence test in a translated Spanish version or in the original English form.

Why a decision between the two versions of the S-B? Could not some other test have been selected? The foregoing questions are asked in lieu of the observation that the S-B is a highly verbal IQ test and does not have separate Performance and Verbal IQ scores, Burns, (1955) has observed the possibility that "verbal tests are unfair to Mexican children..."

Method

Fifty Spanish-American (S-A) children of the 4th grade, drawn from five different schools in Albuquerque, New Mexico, were given Form L of the 1937 S-B. Subsequently, Form L of the same test was administered to these students in a Spanish translation made by Dr. Jose Germain Cebrian of Spain. Four weeks elapsed between testing. All testing was done by a bilingual person who spoke the Spanish native to the area.

The subjects were selected randomly from 5 schools located in different sections of the city. The age of the subjects ranged from 9 years and 3 months to 11 years and 9 months with an average age for the group of 9.85 months. Twenty-two of the subjects were boys and 28 were girls.

Nothing is said in the study concerning the nature of the Spanish translation of the S-B. This is a very important factor to consider for many reasons, only one of which concerns the type of normative population that was used in its standardization, if it was standardized at all. It is quite possible that only the wording was changed.

Results

A correlation of .36 was found between the English and Spanish versions of S-B. The author reports that Terman and Merrill reported a correlation of .93 between form L and Form M of the S-B.

The performance of the subject on the English version of the S-B was significantly better than their performance on the Spanish version of the test. (See Table 1).

Table 1

Range, Mean, SD, and Mean Difference of Intelligence Test Scores

	t value Mean dif.	Form M English Version	Form L Spanish Version
Range		58-121	62-97
Mean	6.85*	66.0	71.8
SD		15.3	9.9

*Significance at .01 level.

Conclusions

The following 5 conclusions were reached concerning the results of the study:

1. The possibility of testing the S-A children with the Spanish version was rejected.
2. The S-B depends to a large extent on education and school achievement and for this reason, bilingual children are able to perform better in the language in which they have formal education.
3. The development of the Spanish language in these children cease when they enter school.
4. Because of the coloquial Spanish spoken by the population in question, the Spanish translation used is not suited to their particular cultural background.
5. The English version of the S-B also present language difficulties and is therefore also an unfair measure for this pop.

Conclusion No. 3 is supported by the findings that the range and variability of the Spanish scores were much smaller than the range and variability of the English scores and a r. of .62 was found between the English test scores and GPA, compared to a r. of 11 found between the Spanish test scores and GPA.

The documentation for the second conclusion came from Mahakian (1939) who reports that English becomes the dominant language for such bilinguals from the 5th grade on.

A general observation concerning the conclusion is that they derive much of their authority not from the data presented but from other sources. It was very difficult to distinguish between the study proper and the review of the literature being conducted throughout the study.

No more information is given in the study concerning how the author obtained GPA than that presented in this critique.

Sample of approach c (comparison of M-As IQ on related verbal and non-verbal tests).

Burns, Pearal Pratt, "The Value of the Leiter Scale in Testing Mexican-American Children." Unpublished Master's Thesis, University of Southern California, 1954.

Purpose

The purpose of this study was to examine the Leiter International Performance Scale (1948) (Revision), and to determine its effectiveness as an instrument to measure the intelligence of Mexican-American children. Of special concern was the problem of screening bilingual children for special classes for the mentally defective. Because of its extensive use in measuring the intellectual ability of M-A children, the Wechsler Intelligence Scale for Children (WISC) was used as a measuring criterion for the Leiter International Performance Scale (LIPS).

Subject

Twenty-five Ss (14 boys and 13 girls) from the Isabell School of Santa Paula used for the study were chosen from two classifications:

(1) Seventh and 8th grade M-A pupils referred by teachers for EMR classes, and

(2) M-A pupils from the 7th & 8th grade considered by teachers to have normal intellectual ability. CA's for the total group ranged from 11-11 to 14-11.

There was no standardized rating scale used to classify the Ss on their socio-econ. background. Quite a bit of description was offered in the text concerning the history and divergent background of the Ss which made the variable seem important enough to control for it in some way.

Method

Both WISC and LIPS were given to all Ss never more than 2 weeks apart. Order of tests altered and time required to administer test was recorded. All but two of Ss were tested during school hours. Careful observation of test administration procedures were observed.

Results

The author was careful to follow the observation by Leiter concerning his test, the LIPS, that the 1948 revision is scaled about 6 months or 5 IQ points too low. Leiter recommended that when comparing the LIPS with other tests whose norm is 100, to add 5 IQ points. The following data was computed with 5 point correction.

TABLE 1

Means, Standard Deviations, and Time Required to Administer the WISC and LIPS

	GIRLS		BOYS		TIME	
	M	SD	M	SD	BOYS	GIRLS
LIPS IQ	89.36	10.10	84.00	17.57	85.0 min.	79.0 min.
WISC: V	84.55	9.50	78.43	6.95		
P	95.82	8.38	95.79	16.33		
FS	89.36	9.31	85.36	11.58	72.8 min.	69.5 min.

For girls r's were: LIPS/WISC p .788, LIPS/WISC V .609, WISC/LIPS .759.

For boys r's were: LIPS/WISC p .723, LIPS/WISC V 5.43, WISC/LIPS .689.

r's for both boys and girls were sig. (.01) on performance and Full Scale and sig. at the .01 level for the Verbal part.

It would have been interesting to have applied tests of significance between means of the different tests and subtests to see if the verbally loaded Full scale and Verbal Scale were significantly different from the performance scale or LIPS.

Conclusions

It was concluded that the WISC and the LIPS tend to measure the same capacities. The findings indicated that the LIPS was a valid test to use with M-A adolescents as the WISC Performance Scale. The same assumption was held for the Verbal and Full Scale WISC IQs.

The case of administration of the LIPS and its non-threatening aspects make it worthwhile for use with adolescent M-As. The culture fairness of the LIPS and its minimum number of timed sub-tests was interpreted as conducive to its usefulness with M-As.

The use of the Verbal and full scale IQs as indicative of the capacity of the M-A have been questioned by Holland (1960), and Altus, (1953).

It would have been of interest if the author had obtained a M & SD for the group of children judged by their teacher as having normal intelligence. The findings may have provided the author with another means of evaluating the LIPS and WISC.

Category II

Study of the M-A via an E Group and Control Group Compared on One or More Tests

There were no direct methodological subdivisions of this category so no artificial divisions are hypothesized. The critiqued studies did tend to separate themselves into somewhat different subgroups, however. The most outstanding characteristic which seemed to differentiate the subgroups in this category was the focus of the study. That is, was the study concerned with focusing its findings on the intelligence of the M-A, the evaluation of the IQ test being used, or the evaluation of both? The first two studies critiqued center their focus on the clarification of the M-A's IQs via the use of the A-A control group and the use of a battery of testing devices. The third study focused on the evaluation of an IQ test via the use of a group of M-A and A-A children, and a fourth study centers on the evaluation of both an IQ test and M-A group by the use of a A-A control group and normative data on the test being evaluated.

Sample of research in category #2 where focus of approach is on M-As IQ

Sivlerstein, A. B.; Shotwell, Anna M.; and Fisher, Gary M., "Intellectual Functioning of the Mentally Retarded." American Journal of Mental Deficiency, 67, pp. 396-401, 1962.

Purpose

The purpose of the study was to investigate the performance of Mexican mental retardates on the Wechsler Adult Intelligence Scale (WAIS). Although the number of Mexican patients available for the study was insufficient to establish norms for their performance, it was thought that an analysis of the scores of even a small number would allow psychologists to objectify some of the allowances they make for cultural factors in the intellectual functioning of the pop. in question.

Fisher (1962) points out the norms of the WAIS might be inadequate for the sample being studied because there was insufficient sampling of subjects of lower intellectual ability in the standardization of the WAIS.

Method

Subjects: The pop. was composed of 53 Mexican and 251 Anglo-American mentally retarded patients at a California state hospital.

Only 5 of the patients were from Mexico but as is the case in many public institutions, the term Mexican was used to describe those of Spanish surname.

TABLE 1

Comparison of Mexican Patients and Controls on Three Demographic Variables

Variable	Mexican	Control	t.
CA	27.7	33.5	3.55 sig. .001
Hospitalization	9.3	13.8	3.16 sig. .01
Social Class	6.3	5.3	4.52 sig. .001

No sig. sex differences found; A total of 56.2% of the sample were male. Social class was determined on the basis of parental occupation (Hollingshead).

Information given in table is all there was available concerning CA. Nothing is said concerning age range in the text and no SD was given.

Procedure: The design called for a comparison of the Mexican patients and the controls

on the WAIS Verbal, Performance, and Full Scale IQs; on the 11 subtest scores; and on three factor scores derived from Cohen's (1957) factor analyses of the WAIS standardization data.

A statistical method was employed to control for possible contaminating variables. Point-biserial correlations were first calculated between the Mexican control dichotomy and each of the WAIS variables, and the corresponding partial correlations were calculated, holding constant chronological age, sex, length of hospitalization, and social class.

Results

The findings (see Table 2) demonstrated that the Mexican patients and the controls did not differ significantly on the Verbal or Full Scale IQs, but the mean Performance IQ of the Mexicans was significantly higher than that of the controls.

Table 2
Comparison of Mexican Patients and Controls on WAIS IQs

Variable	Mexican	Control	r	Partial r
Verbal IQ	61.7	62.4	-.03	-.03
Per. IQ	68.2	62.5	.13	.18*
Full IQ	62.4	60.3	.08	.09

*Significant at .01 level.

The discrepancy between the mean Verbal Performance IQs was completely negligible for the controls, whereas for the Mexican patients, this discrepancy was highly significant in favor of the Performance IQ.

In the comparison of the Mexican patients and the controls on the 11 subtests of the WAIS on the Verbal Scale, the controls obtained significantly higher mean scores than the Mexican patients on Information and Vocabulary. On the performance subtests, the Mexican patients obtained significantly higher mean scores than the controls on Digit Symbol, block Design, and Object Assembly.

The Mexican patients and the controls did not differ significantly on Cohen's Verbal Comprehension and Memory factors, but the mean score of the Mexican patients on the Perceptual Organization factor was significantly higher than that of the controls. (see Table 3).

Cohen concluded that the Verbal and Performance IQs do not represent the actual functional unities in WAIS performance, instead he identified 3 other factors: Verbal comprehension, Perceptual organization and Memory (arithmetic and digit span).

The statistical approach was used by the authors instead of the traditional approach of employing matching for the variables described because of the methodological problems involved in matching. Briefly, matching may, "(a) restrict the populations to which conclusions may be generalized, (b) bias the conclusions in favor of one or the other group, and (c) result in a loss of statistical power if the matching is not explicitly taken into account in tests of significance."

No statistical information is given concerning the comparison of mean Verbal and Performance IQs of the Mexican patients.

Table 3

Comparison of MExican Patients and Controls
on WAIS Factor Scores

Variable	Mexican	Control	r	Partial r
Verbal Comp.	3.6	3.9	-.09	-.07
Perceptual Org.	5.6	4.0	.26*	.22*
Memory	3.0	2.8	.04	.04

*Significant at .0001 level

Conclusion

The findings suggest that the greatest difference between Mexican mental retardates and controls lies in the factor of Perceptual organization. Since the most significant of the difference observed, on Object assembly, was just two Scale Score points, and accounted for less than 10% of the variance, the results do not provide clinical psychologists with much help in objectifying the allowances that they make for cultural factors in intellectual functioning. If anything, they suggest that these allowances should be made with caution lest more credit be allowed the patients than is due them.

Although there may be some criticism of the study concerning the inadequacy of the norms used in validating the testing instrument, the study was well conducted. The conclusions tend to run contrary to those in most of the literature in this area.

Sample of research in category #2 where focus of approach is in on M-As IQ.

Carlson, Hilding B. and Henderson, Norman, "The Intelligence of American Children of Mexican Parentage." Journal of Abnormal and Social Psychology, 45, pp. 544-551, 1950.

Purpose

The purpose of the study was to ascertain whether any group differences in intellegence between white children of non-Mexican parentage (A-A) and those of Mexican parentage remain when research conditions are controlled as much as possible.

A control group of A-A children were compared in intelligence over a 5½ year period with an experimental group of children whose parents had been born in Mexico. An attempt was made to control all known variables that have been demonstrated to be of significance in similar studies. These important variables were: (1) rural vs. urban environment, (2) general socio-economic level, (3) total cultural complex, (4) amount and quality of formal education of Ss parents, (5) effects of inadequate diet, (6) prejudice on part of the examiner, (7) motivation, and (8) bilingualism.

Method

Control for the important variables was attempted as follows:

(1) Rural vs. urban environment-All of the Ss lived in Los Angeles and all had been born in the U. S. and had attended no other schools other than those of a restricted geographical area of the city.

(2) Socio-econ. level and total cultural complex--Only those children who lived in a fairly homogenous, restricted and older section of Los Angeles were used.

(3) Education - Ss in both the control and experimental groups attended the same school and were taught by the same teachers.

Number 1 and 2 controls are held in question because of the tendency of the M-A to group and form ghettos. In these ghettos, the traditional handicapping factors may continue to operate. Although no information is given concerning the cultural grouping of individuals in the area, the ghetto situation is highly probable in Los Angeles.

If the M-A enters the school handicapped, he cannot take equal advantage of educational opportunities. (Jones, 1954).

(4) Health factors - All pupils who were absent 25% or more of the time were eliminated from the study.

(5) Prejudice - The tests were not administered by the authors. Instead, the recorded routine test results obtained by many duly appointed school psychological test examiners were used.

(6) Motivation - All Ss were tested under the same general physical conditions. Since there was no segregation of Ss into control and experimental groups and all were tested in like manner, at least the external factors in the motivation of subject in both groups were controlled.

(7) Bilingualism - This factor could not be controlled. If it had been, the number of cases would have been too small.

It was thought that more convincing evidence could be had if the same children were involved at a number of age levels, constituting what might be termed a longitudinal approach.

Subjects: The selection of children for the experimental group were those who met the methodological approach and the following two criteria: (a) that the students had taken not less than three tests of intelligence, including the California Test of Mental Maturity (CAT) and (b) that the students were in the 5th or 6th grades during May and June of 1947. Total number of experimental Ss was 115.

The control group (N=105) were selected from the same geographical area, and met the same methodological control conditions. They differed only in that they were A-A Negroes and Orientals were not included in the control group.

Mean CA of M-A tended to be somewhat higher than that of the A-As, (See Table 1).

Intelligence tests: IQ scores available for the study were: Detroit Beginning First Grade (DBFG), Detroit Primary (DP), Pintner-Cunningham (P-C), Kuhlmann-Anderson (K-A), CTMM (Elementary short-form), and the Stanford Binet (SB).

Because of a communication problem with bilingual students, the problem many times is not one of prejudice but one of communication.

The problem of motivation is not a superficial one being solved by superficial means. It is much more deep rooted and perhaps finds its beginnings in childhood revolving around such factors as Achievement Training (Rosen, 1959), Delay of Gratification (Mischel, 1961), and other factors which permeate the individual's achievement under all academic situations. Equating testing conditions to the concept of motivation does not guarantee similar motivational behavior.

Results

For the experimental group, the difference in mean IQ between the first and third testing period is statistically significant. For the control group, only the difference between the first and second testing period is statistically significant. In every instance, the mean IQ of the Experimental group is lower than the corresponding mean IQ for the control group, critical ratios indicating that the differences at each testing period are not considered to be due to chance

No level of sig. was included.

TABLE 1

Mean IQ Scores for Experimental and Control Groups

	TEST USED	Experimental Group			Control Group		
		N	M Age	M IQ	N	M Age	M IQ
First Testing	Detroit First Grade	97			77		
	Pintner-Cunningham	16			25		
	Detroit Primary	2	79.1	94.1	1	75.9	105.0
	Not Indicated				2		
Second Testing	Detroit Primary	65			77		
	Cuhlmann-Anderson	17			19		
	Pintner-Cunningham	14			3		
	Detroit First Grade	11	106.9	91.0	2	103.6	99.6
	CTMM	3					
	Terman-Merrill, 1937 Revision of Binet	5					
	Not indicated						
Third Testing	CTMM	115	137.0		105	127.6	
	Language			89.6			103.5
	Non-Language			89.3			103.8
	Total			88.5			103.3

No significant dif. were found between language and non-language scores on the CTMM.

It might have been helpful if range of SD had been given.

In an effort to control for possible spurious results, smaller groups of experimental and control subjects were isolated, all of whom had taken the same tests at each testing period. A total of 52 subjects composed the experimental group and 65 subjects composed the control group. The findings of the substudy designed to test for possible similar results were similar as these in the study proper; differences between the 1st and 2nd testing periods were significant for both groups and there were significant differences between groups at each testing period. (See Table 2 for findings.)

TABLE 2

Mean IQ Scores for the Sub-Experimental and Sub-Control groups

TEST USED	SUB-EXPERMENTAL GROUP			SUB-CONTROL GROUP		
	N	M Age	M IQ	N	M Age	M IQ
Test 1: Detroit First Grade	52	76.8	97.00	65	74.4	107.71
Test 2: Detroit Primary	52	107.2	91.46	65	102.1	101.16
Test 3: CTMM	52	134.1	91.44	65	127.3	105.55

In an effort to clarify the predictive stability of the testing devices over a period of the time inter-correlations were run between testing periods. Pearson-product-moment rs reveal higher rs between the testing periods for the control group than for the experimental group. When subgroups were used the rs between testing periods became similar in magnitude for the two groups. (Table 3)

The test reliability correlation coefficient computed for the study proper are sufficiently low to cast some doubt on the reliability of the tests for both Groups but particularly for the experimental group.

TABLE 3

Pearson Product-Moment Correlation Coefficients
Between Testing Periods

Between Testing Periods	Experimental Group M-A	Control Group A-A	Sub-Experimental Group M-A	Sub-Control Group A-A
1-2	.35	.58	.61	.61
2-3	.49	.62	.79	.60
1-3	.38	.45	.48	.48

Discussion

The experimental group had a lower mean IQ at every testing period than either the control group or the groups used in the standardization of the tests involved in this study. The results, are consistant with other findings in showing children of Mexican parentage to have lower IQ score than A-A even though attempts were made to control for defects of previous expermentation. Also consistent with previous findings is the observation that differences in IQ between M-A and A-A children increase from the first testing period at about six and one-half years of age to the last testing period at about eleven years of age. This difference is primarily due to the drop of IQ in the M-A group.

Uncontrolled factors which may have accounted for the results were: (1) rural urban parental background (2) possibly a more limited vocab. not only in English but also in Spanish for M-A children, and (3) possibility of differences in motivation at time of testing.

Sample of research in catagory #2 where focus of approach is on the IQ test

Stablin, John E.; Willey, Darrell S.; and Thompson, Calvin W., "An Evaluation of the Davis-Eells (Culture-Fair) Test Using Spanish and Anglo-American Children." Journal of Educational Sociology, 35, pp. 73-78, 1961.

Purpose

The present study was designed to evaluate the culture-fairness of the Davis-Eells test of General Intelligence, or problem Solving Ability (DET). This was done by comparing the performances of two groups of culturally different elementary school children on the DET with their performances on three other measures. The other measures were: academic achievement, Metropolitan Achievement Test Battery (MATB); a group test of intelligence, Primary Mental Abilities Test (PMA); and a vocabulary test, composed of 50 words selected for each grade level according to three Thorndike, Gates, and Dolch word frequency lists (VOC). In addition, a modified form of the Sims Socio-economic Score Card (SI) is used to obtain an estimate of each child's socio-economic level. The assumption was made that the two culturally different groups of children should perform equally well on the DET if it is truly a culture-fair test and there are no basic herditary differences present.

Method

Subjects: Eighty-three Anglo-American (A-A) children with a mean CA of 9.4 years and 127 Spanish-American (S-A) children with a mean CA of 10.2 years constituted the pop. for the study. These subjects were enrolled in grades 2 through 5 in a "typical" southern New Mexico public school with approximately an equal number of subjects in each grade. The difference in mean age levels was due to policy of the state of New Mexico to place S-A children in pre-first.

It is not clear just what procedure was used in administering the SI. The procedure for taking this instrument requires that the person make decisions on 42 occupational titles. For each given occupation, the subject indicates whether he feels that the people in that occupation belong in the same social class as himself and his family, or in a higher or lower class. Anastasi (1963) Since no reference or explanation concerning the nature of the instrument used is given there is some doubt concerning the usefulness of such a complex scale with 9 year old children.

Procedure: The DET,VOC, andSI were administered during the Fall term. IQs from PKA and achievement ratios from the MATB were taken from the cum. record files. Both tests had been given as part of the regular school program the prior spring. The VOC Test was administered orally subsequent to the SI interview.

It appears as if the SI might have been administered orally to the children although all that was said about it in the study proper is included here.

Results

Table 1 summarizes the comparisons of the mean scores obtained for the two groups on the five measures. The mean scores for the two groups differ on all five variables, each significant beyond the .01 level. The distributions of scores were reasonably normal for both groups on all measures except the vocabulary test which was highly skewed due to the nature of the test which was based on words common to each respective grade level.

TABLE 1

Comparison of the Means for the Two Groups on the Five Measures

	Anglo-Americans			Spanish-Americans			
	N	M	-	N	M	-	t
DET	83	104.1	12.9	127	85.6	13.3	10.0
PKA	63	111.5	11.1	104	89.8	12.5	12.2
MATB	57	126.4	21.0	102	108.9	27.1	4.2
VOC	83	46.0	3.5	127	34.4	9.2	11.4
SI	83	38.8	3.9	127	25.5	6.7	5.2

The differences between r's were tested for significance and with the exception of two, the DET/PKA and PKA/ VOC failed to meet the .05 level of confidence. (see Table 2)

Although there was only a four year span in grade level r's might have been altered by controlling for the spurious effect of age.

TABLE 2

Correlations Among the Five Variables

	A-A		S-A		t
	N	r	N	r	
DET MAT	57	.28	102	.31	0.2
DET PMA	63	.43	114	.65	2.9*
DET VOC	83	.25	127	.46	1.7
DET SI	83	.11	127	.30	1.4
PMA MAT	53	.56	98	.68	1.1
PMA VOC	63	.26	98	.68	4.3*
PMA SI	63	.34	114	.27	0.5
MAT VOC	57	.28	102	.38	0.7
MAT SI	57	.26	102	.33	0.5
VOC SI	83	.23	127	.35	0.9

*Significant at .05 level

For purposes of comparison, only the DET and PMA IQ's were used (See Table 3). There was no apparent age change on either the PMA or the DET.

TABLE 3

Comparison of Four Grades

		A-A		S-A	
Grade		DET	PMA	DET	PMA
2	M	96.8	222.3	84.5	94.1
		11.4	10.8	12.4	12.0
	N	20	17	37	36
3	M	110.7	113.9	87.5	91.5
		9.9	10.4	14.7	11.1
	N	23	18	23	21
4	M	100.9	109.9	83.11	82.5
		13.2	10.4	13.8	11.5
	N	20	16	28	27
5	M	107.2	106.8	88.2	88.7
		12.6	12.7	12.4	11.3
	N	20	12	39	30
Total	M	104.1	110.9	85.6	89.8
		12.9	11.1	13.3	12.5
	N	83	63	127	104

Tireman (1943) using a longitudinal approach over more grades has found a gradual and significant deterioration in the measured intellectual functioning of the lower Class S-A.

Although this was not a longitudinal study, it might have been expected that there might have been some deterioration demonstrated in Table 3.

Discussion

The DET failed to give a measure with less difference between the groups than other purportedly culturally loaded, measure of intelligence and achievement. The DET thus failed to provide a performance measure less influenced by cultural experience factors than other commonly used tests.

The findings for the DET by the author are very similar to those reported by Rosenblum, et. al. (1955). The findings by the present author and Rosenblum run contrary to the historical trends described by Burns (1955) in the summary of the literature, "A survey of the literature..., reveals an evolution of thought regarding the abilities of these subjects, away from the inferior rating given them in verbal group tests, to the normal ability obtained in non-language and culture fair tests." (underling added by UHP)

Sample of research in category #2 where focus of the approach is on the evaluation of both M-A IQ and IQ test

Altus, Grace T., "W.I.S.C. Patterns of A Selective Sample of Bilingual School Children," The Journal of Genetic Psychology, 83, pp. 241-248; 1953

Purpose

The purpose of this study was to attempt a comparison of the intelligence test patterning of a selective portion of the M-A bilingual minority to that of a unilingual A-A sample equated on the basis of age, sex, and performance I.Q.

The Wechsler Intelligence Scale for Children (WISC) was chosen for the investigation because it contains both a Verbal and a Performance IQ as standardized on the same pop.

Method

Ss were enrolled in the years 1949-50 and 1950-51 at the Santa Barbara County schools in California. Most of the Ss had been referred to the Guidance Dept. for IQ tests relative to preliminary screening for special classes for the mentally retarded. A dozen members of the bilingual sample, older and brighter than the former, were tested as a part of another research surveying reading.

A child was considered bilingual if his cumulative record indicated that Spanish was spoken exclusively, or that both Spanish and English were spoken, in the home.

The majority of the unilingual group came from the same two sources. Because of the method of selection, both samples are positively skewed and are not assumed to be representative of the total school population from which they were derived.

The WISC was given to all Ss by the author. Description of Groups and WISC performance IQ given in Table 1. Since the mean age was close to 11 years for both groups over-agedness is the case for both groups in grade placement.

If the "dozen" members of the splinter group included in the test pop. were brighter and not referred for placement, it might be expected that their test patterning might be different than that of the MR referals. It is very difficult to understand just what sources the unilingual group came from. What does the author mean by "the majority?"

TABLE 1

Description of Groups

	Unilinguals	Bilinguals	Diff.	t
Age in years	10.98	11.14	.16	.30
SD	3.05	2.64		
Sex ratio (M:F)	27:25	35:32		
WISC Perf. IQ	86.43	84.01	2.42	.88
SD	14.40	15.75		

TABLE 2

Mean IQ's and Subtest Scaled Scores on the WISC and Their Differences for Unilingual and Bilingual Groups

	Unilinguals		Bilinguals			
	Mean	SD	Mean	SD	Diff.+	t*
Verbal IQ	88.98	14.20	72.07	14.25	16.91	6.43*
Performance IQ	86.43	14.40	84.01	15.75	2.42	.88
Full Scale IQ	86.89	14.02	75.13	14.58	11.76	4.49*
Inf.	7.90	2.95	4.94	2.66	2.96	5.69*
Comp.	8.81	2.59	6.18	2.69	2.63	5.48*
Arith.	7.89	2.98	6.24	2.70	1.65	3.11*
Sim.	8.43	2.76	5.11	2.88	3.32	6.38*
Voc.	8.47	2.71	4.65	2.48	3.82	7.80*
P.C.	9.17	2.44	7.62	3.38	1.55	2.92*
P.A.	7.74	2.88	7.28	2.98	.46	.85
B.D.	7.74	2.72	7.46	2.43	.28	.58
O.A.	7.61	2.23	8.07	2.65	-.46	1.02
C.	8.10	3.06	8.05	2.95	.05	.09

*Differences significant at or beyond .01 level.
+Unilingual mean minus bilingual mean.

Results

While the difference in Perf. IQ between the two groups had been shown to be slight and insignificant, the average difference in Verbal IQ's was nearly 17 points and highly significant, favoring the unilingual group.

The English-speaking group followed Seashore's description in terms of a close correspondence between Verbal and Perf. IQ's.

The only subtest showing a significant difference between bilingual and unilinguals on the Perf. Scale is Picture Completion.

Only six of the bilingual group had higher verbal than Perf. IQ's whereas an even half of the unilingual group were in this category.

Conclusion

A unique subtest pattern emerged for the bilingual group, which was dissimilar to the customary adult WISC pattern for the mentally retarded. The pattern for the unilingual group was very similar to the pattern described by Seashore.

The author is not very clear about how the unilingual or bilingual subtest patterns compare or do not compare to Seashores description. Since the observation concerning the differences in sub-test patterns is one of his most important findings, it would seem there would have been some graphic or numerical representation of the finding.

Much of the study seemed to concern itself with the discussion of observations concerning the findings which were admittedly insufficient to draw the conclusions from, yet these were widely discussed.

Category 111

Study of the M-A via 4 Groups with the Critical Evaluation Centered on the IQ tests

In this category the focus seemed to be the critical evaluation of the IQ tests being used for the development of better measures of the intelligence of M-A children. The first study discussed represents an approach that is widely used in evaluating the IQ tests used with M-A children. An IQ test is evaluated against 4 groups which have been dichotomized by the use of other instruments and/or the judgement of experts. Complex statistical methods are used to evaluate not only the test but separate items within the test for their usefulness in measuring the IQs of the M-As. The second study was chosen for its originality in approaching the assessment of the M-A intelligence. It shows not only a sophistication of statistical treatment but originality in its measurement of intelligence.

Sample of Research in category #3 (evaluation of Usefulness of IQ test for M-A population via 4 groups)

Kidd, Aline M., "The Culture-Fair Aspect of Cattell's Test of g: Culture-Free." Journal of Genetic Psychology, 101, pp. 343-364, 1962.

Procedure

The purpose of the study was to: (a) investigate the culture-fair characteristic of the Cattell Culture Fair Intelligence Test; (b) determine which items are more culture-fair; and (c) discover which aspect of intelligence actually are measured by the test.

Method

Warner's Index of Status Characteristics based upon occupation, source of income, quality of the home, and quality of the dwelling area was applied to select four groups of children between the CA of 10-0 and 11-0 from public, private and parochial schools in Tucson, Arizona. The sample included:

25 Upper socio-econ. status A-As
(13 male, 12 female)
25 upper socio-econ. status M-As
(12 male, 13 female)
25 lower socio-econ. status A-As
(12 male, 13 female)
25 lower socio-econ. status M-As
(13 male, 12 female)

The number of male and female Ss were nearly equated to counterbalance the linguistic superiority of girls and math and mechanical superiority of boys.

The revised Stanford-Binet, Form L and Test of g: Culture-Free were administered to each child. To balance fatigue, approximately one-half of the girls and half of the boys in each group were given Binet first and then test of g, with remaining children taking the tests in the reverse order.

Results

Set of results, 1,: For the study pop. the mean IQ for Test of g was 93.88, SD = 19.85 (Cattell reports: M=100.0,

Another study critiqued in this paper (Jensen, 1962) concluded that it would be difficult to find a sufficient number of M-As with high IQs. A possible reason for his failure could have been that he did not include parochial schools in his sample. Parochial schools draw a selected sample of M-As from the public schools. This sample is heavily loaded with some of the more middle class-oriented type M-As which are more likely to do better on middle class-oriented type tests. (Holland, 1960).

SD=24.4 for standardizing pop.) The mean for the Binet was 104.7 and the SD was 18.43 (at 10 year old level Binet mean is 96.7, SD=16.5).

Set of Results, 2: Bartlett's test for homogeneity of variance indicated that the four groups were not heterogenous in variance and may be considered to be random samples from groups with common variance.

Distribution of the IQs obtained from the Binet and from the test of g approximated normality.

An analysis of variance revealed that the dif. between the national and between the econ. groups were sig. (.01). (See Table 1.) The dif. shown in interaction between national groups and sexes was sig. (.01). A t test revealed that the source of dif. was between the means of the lower class male A-As and the lower class M-As ($t=3.085$, $p<.01$). A dif. between the means of lower class female A-As and of lower class female M-As was sig. at the 7% level ($t=2.025$).

TABLE 1
Analysis of Variance of Binet IQs

	Source	df	ESS	MSS	F
National groups	A	1	4998.49	4998.49	16.93*
Economic levels	B	1	6905.61	6905.61	23.39*
Sex	C	1	7.25	7.25	.03
Interactions	AB	1	193.21	193.21	.65
	AC	1	2991.38	2991.38	7.76*
	BC	1	178.37	178.37	.60
	ABC	1	606.08	606.08	2.05
Error term		92	27165.40	295.28	
	TSS	99	42345.79	4274.40	

*.01

An analysis of variance was also run on the Test of g and the two tests combined, using normalized T scores to correct for the different Ms and SD of the two tests.

In the analysis of the Test of g the dif. between national groups was sig. (.01). The dif. between the econ. groups was sig. at the .05 level. The interaction between econ. and sex and the interaction among national groups, econ. level, and

sex were both sig. (.01). Only the dif. between Ms of lower class A-A males and lower class M-A males was found to be sig. (t=2.672) when a t test was run to find the source of dif.

A summary of the analysis of the combined tests is as follows:

Sig. dif. were found in interaction between (b) national groups, econ. levels, and sex; (c) econ. levels, sex, and tests; and (d) national group, econ. levels, sex, & tests. Sig. for above was at .02 level.

In determining the sources of the dif. it was found that the t for the dif. between the means obtained: (a) from the lower class A-A female and lower class M-A females was sig. at the 4% level; (b) from upper class A-A females and upper class M-A females was sig. at the 3% level; (c) from lower class A-A males and lower class M-A males was sig. beyond the 1% level.

Rhos were run to show the relationship between IQs obtained from the two tests. All Rhos were sig. at or beyond the 1% level. (See below.)

Group	Rho
Upper class A-A	.662
Upper class M-A	.745
Lower class A-A	.636
Lower class M-A	.682
Total.	.681

Set of Results, 3.: Because the analysis of variance showed no sig. dif. for the interaction of national groups and tests, it was apparent that there were some items on both the Binet and Test of g which did not differentiate between the two ethnic groups. Item analysis was done to locate these items. The sample was divided into groups A and B and randomization was followed so that representative samples of males and females, of upper and lower socio-econ. status children, and of A-As and M-As were maintained in each group.

The Chi-square test was applied to the data from Group A for the entire range of Binet items administered except 3 items that were passed or failed by 90% of the group. Only three items were found not to

Information for the sig. of interaction in the combined tests not included in text of study proper but can be found in Table 3, p. 348.

differentiate between the ethnic group at or beyond the 10% level. The three items discovered in group A were re-tested in group B and in the total group with the same results for group A.

Similar Chi-square tests were applied to the items on the Test of g. There were 24 items which did not differentiate between the A-As and M-As at or beyond the 10% level of confidence. The items were then studied in group B and 2 items were found to discriminate and were subsequently dropped.

Group B A-As obtained a mean score of 12.51 for the 22 items and Group B M-As obtained a mean of 11.51. The t for the dif. between means was .888 which reaches only the 55% level of sig.

Set of Results, 4.: A factor analysis of the culture-fair items from the Test of g indicated that four reference factors were operating: The complete centroid method of factor extraction was used, both the Tucker and the Guilford-Lacy criteria of completeness of extraction.

a. Ability to overcome inappropriate set.

b. Localization, or ability to localize the element of an item.

c. Magnitude, or awareness of size concepts.

d. Progression, or ability to understand the essential change in a progressive series.

Set of Results, 5.: A second-order factor analysis indicated that two general reference vectors were operating:

a. Recognition of Total Figure-ground Relationships, or awareness of all interrelationships.

b. Set, or tendency to note one obvious element to the exclusion of other important elements.

6. Both the first-order and the second-order reference vectors appeared to operate similarly, if not identically, within all of the groups used in this study.

Conclusions

An analysis of variances of the total scores of the Test of g indicates that some items of the test are not culture-fair in that the difference between national and econ. groups, as well as the interaction between econ. level and sex, are sig. The t test shows the greatest dif. to be between the lower class M-A male and the lower class A-A males. Some of the results of the three analyses of variance made may have been influenced by two uncontrolled variables. The length of residence in the M.S. by the families of the M-A children was not taken into account, nor was the anthropologically demonstrated fact that upper econ. status individuals tend to be culturally assimilated more rapidly than do lower class individuals.

The culture fair items isolated by the author could provide very definite guide lines for the development of instruments designed to get valid measurements of the M-A intellectual ability.

PERCEPTUAL DEVELOPMENT AND CONCEPT FORMATION

Jensen, Arthur R., Learning Abilities in Mexican-American and Anglo-American Children. California Journal of Educational Research, 12, pp. 147-159, 1961.

Purpose

The general purpose of this paper was to investigate a method for assessing the educational potentialities of Mexican-American (M-A) children, as well as those of other national subculture and ethnic groups, by using tests which provide direct measurements of their present learning ability. Specifically, the question with which the study was concerned was whether or not groups of M-A and Anglo-American (A-A) children equated in IQ, as measured by the California Test of Mental Maturity (CTMM), are also equal in learning ability, as measured by certain standard learning tasks.

Method

Learning Tasks:

The learning materials were of two types:

1. Familiar. There were two equivalent forms of the test using the familiar materials, which consisted of 12 common, readily recognized objects such as a water glass, a bar of soap, a key, etc.

2. Abstract. These materials consisted of seven plastic forms approximately $1\frac{1}{4}$ inches in diameter and $\frac{1}{2}$ inch thick, and of the following shapes and colors: yellow diamond, green diamond, yellow triangle, etc.

There were three types of learning tasks:

1. Immediate Recall. This test always preceded the other tasks. Its main purpose was to familiarize the subjects (Ss) with the materials. Sitting opposite from the S the examiner (E) would withdraw from a box either Familiar or Abstract materials and place them in front of the S asking him to name them. The S could name them any way he wanted. After all the 12 objects had been demonstrated and named, the E would ask the S to name as many as

This is a very well thought out procedure. In this fashion the child could become acquainted with all objects early in the testing experience.

he could. As the S named the materials they were placed in front of him. When the S could not recall the rest after 30 seconds of deliberation they would be brought out and renamed. Then the whole procedure was repeated until all objects were recalled. Prior to the materials being put in the box after the first naming, the S was given 10 seconds to look the objects over. S's score was the total number of uncalled items before attaining criterion.

2. Serial learning. After the Recall test, the E placed a set of 12 (7 in the Abstract test) inverted, plain white cardboard boxes (4" x 4" x 8") in a row before the S, who was turned around so he couldn't see. S had to learn which object was under each box, by starting at his left end of the series, guessing what was under the box to see if he was correct. The subject repeated this until he had guessed all of the items correctly. His score was the total amount of errors made before all guesses were correct.

3. This test made use of both forms of the familiar objects. One set of 12 objects was fastened to the outside of each cardboard box, with the other set placed under the boxes. The S's task was to learn what was inside each box. The order of the boxes was completely re-arranged on each trial to rule out serial learning. The Paired-Associates task was preceded by the Recall test for the objects placed inside the boxes. S's score was the total number of errors made before attaining the criterion of correctly guessing the objects under all 12 boxes.

Subjects:

The population used was selected from the 4th and 6th grades in 5 public schools in Contra Costa County. The schools drew mainly from a semi-rural, laboring class population. All Ss were native born Americans and were from the lower socio-economic strata of society (as judged by the father's occupation, locality of the residence and the teacher's knowledge of the child's home). The M-A group was bilingual. The A-A group came from monolingual English speaking

This is a rather complete sample description. The author lacked only some information concerning the extent of bilingualism in the M-A children and a definition of the M-A to have been complete.

families. All Ss had both parents living in the home and were free of physical defects.

While A-A Ss were selected at random within conditions called for by experimental design, it was very difficult to find enough M-As to fill one of the cells in the experimental design. In order to fill the cell requiring 9 M-As with IQs over 110, it was necessary to obtain Ss from a large school in another county. The author concluded that it would have been virtually impossible to conduct a large scale study of this type if the experimental design had called for the matching of M-A with A-A along the entire range of IQs.

While the author's observation concerning the feasibility of M-As and A-As along the entire range of IQs may be correct for his geographical area, (later on he describes a class as typical which has 6 M-As to 22 A-As) it may be an overgeneralization for him to state that the same situation would be true in other places. For example, in areas in southern California there are many middle class M-A children attending private schools where the upper levels of the IQ range might be adequately represented.

No background description is given for the 4 M-As brought in from the other school to fill the cell.

Experiments:

Three experiments were conducted.

1. The purpose of this experiment was to compare the learning scores on both the Familiar and Abstract tasks for Recall and Serial Learning in groups of "Bright" and "Dull" M-As and A-As. The Dull groups had CTMM IQs ranging from 73 to 89. The bright group had IQs in the range of 111 to 126. Table I shows the mean IQs and ages in each group. All subjects were selected from the 4th grade and ranged in age from 9 to 11.

Two Es gave the tests to the two groups of Ss. The tests were administered individually in a single session, which took, at an average, about an hour.

Table I

Mean Age and IQ of Sample Groups

Group	N	Mean Age	Mean IQ	SD
Bright Mexican-American.....	9	9.33	117.33	6.10
Bright Anglo-American.....	9	9.11	115.67	4.27
Dull Mexican-American.....	9	9.67	82.89	5.82
Dull Anglo-American.....	9	9.78	81.78	3.93

2. The IQ range in this experiment ranged from 60 to 120 or above. Because of scarcity of high IQ M-As, Ss from the sixth grade were used for whom recent CTMM results were available. The same conditions for selection mentioned earlier were applied. There were two M-As and two A-As in each of the four IQ levels: 60-70, 80-90, 100-110 and 120 or above. The 16 Ss were tested in Recall, Serial Learning, and Paired-Associates tasks, using only Familiar objects.

3. The purpose of this experiment was to determine the equivalent forms reliability of Recall and Serial tasks using the Familiar objects. One E administered to every child in a single 4th grade class (N=22) both the Recall and Serial tests, using Forms A and B two weeks apart. The class was typical for the school district; 6 of the 22 Ss were M-As, all considerably below the class average in IQ.

Limits on this study by the geographical area in which it was conducted are again evident. The very small population used in this second experiment was later to influence the findings which might have been favorable to the outcome of the study. Duplicating this study in an area in which sufficient population would be available would be very informative.

Results

Experiment 1:

Analysis revealed no main effect for "nationality". However, a ($p<.05$) significance was found for "nationality" x IQ interaction, with Dull M-A and Dull A-A groups, widely separated in learning scores, and Bright groups more similar. The Dull M-A group did about as well as both of the Bright groups. Only the Recall tests showed a significant overall difference ($p<.05$) between IQ levels.

The r between the Recall and Serial Learning tasks based on the composite Abstract and Familiar scores of all Ss was .64, ($p<.01$). The r between the learning of Abstract and Familiar objects (based on the composite Recall and Serial scores was .68, ($p<.01$). The r's were interpreted as rather substantial reliability of these tests.

Experiment 2:

The mean number of errors on each test for all Ss was: Recall, 7.5; Serial Learnings, 26.1; Paired-Associates learning, 73.5. The difficulty of Paired-Associates learning over Serial Learning was questioned by the authors because both had 12 facts or S-R connections to be learned.

Analysis of variance revealed that with the small number used in this experiment only the Serial Learning test achieved significant differences. The other tests were consistent with the Serial Learning test (as would be expected with an intra-class of .75 between the tests). The Serial test showed significant differences between IQ levels, and showed a significant "nationality" x IQ interaction. These findings are similar to those shown graphically in Figures 1 through 4.

Experiment 3:

Because of the promise demonstrated by the *Serial Learning* test, reliability figures were obtained for it by using the equivalent form method after a two week waiting period. Reliability was computed by the Hoyt Method, using analysis of variance. For the Serial Learning of Familiar objects, the reliability of composite scores based on the two forms of the test was .88; the reliability of scores on a single form was .81.

The r between IQ (CTMM) and Serial Learning scores in this class is only -.30, which is not significant for this small sample.

Figure 1

Mean error scores of Dull (IQs 73-89) and Bright (IQs 111-126) Mexican-American and Anglo-American fourth grade children on the Recall test for Familiar Objects.

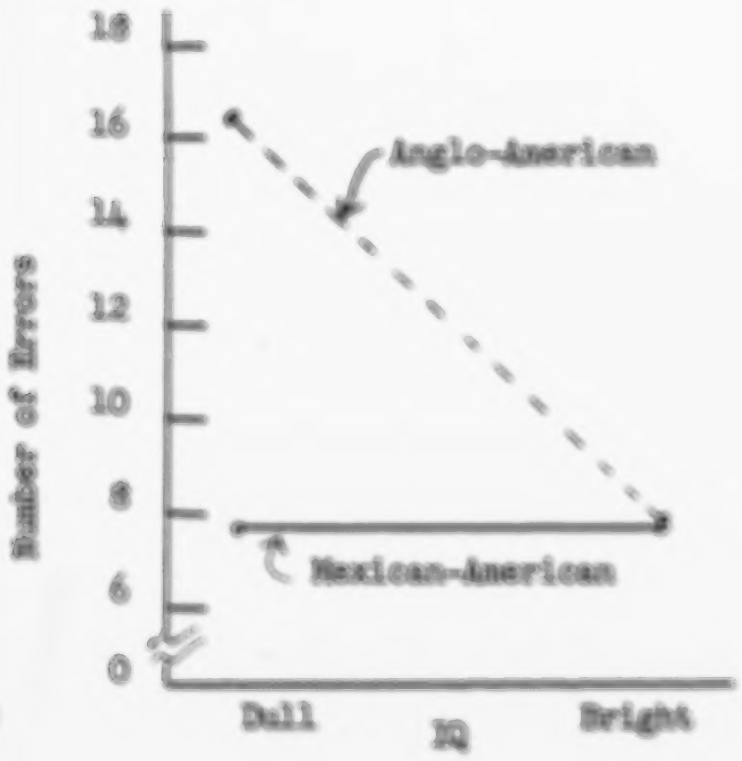


Figure 2

Mean error scores of Dull and Bright Mexican-American and Anglo-American children on the Serial Learning of Familiar Objects.

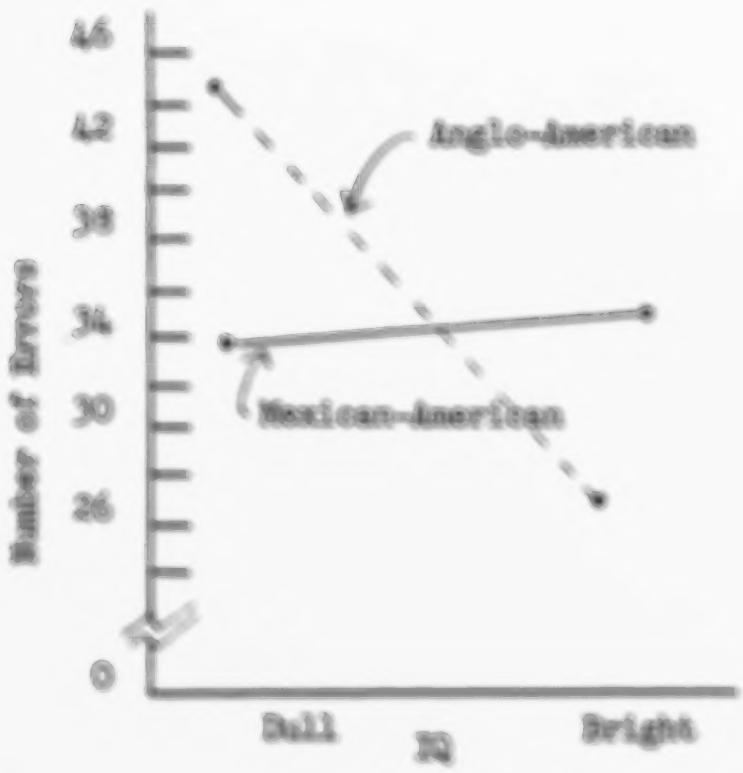


Figure 3

Mean error scores of Dull and Bright Mexican-American and Anglo-American children on the Recall test for Abstract objects.

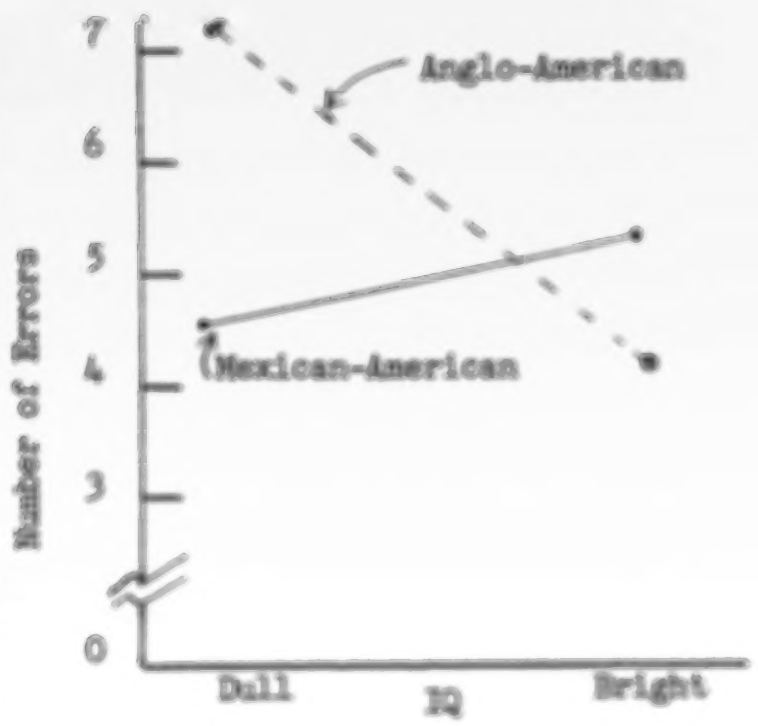
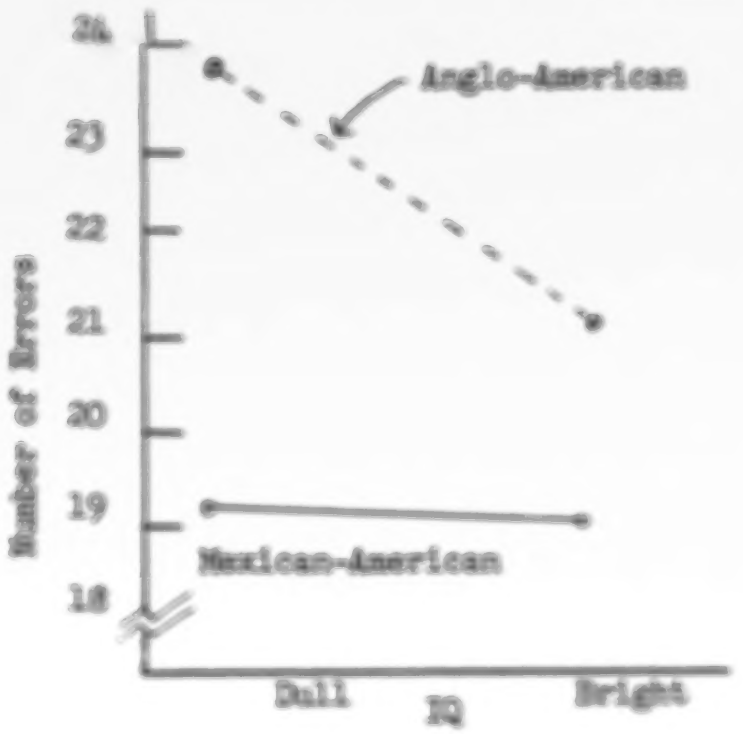


Figure 4

Mean error scores of Dull and Bright Mexican-American and Anglo-American children on the Serial Learning of Abstract objects.



Conclusions

On learning tasks used by this study M-A children with low IQs performed significantly better than A-A children with low IQs. Dull M-As performed as well as Bright A-As and M-As. The CTMM seems to have been successful in discriminating between children of high and low IQ in the A-A group, but did not do so with the M-A group.

These are impressive, well-backed findings which not only clarify the inadequacy of IQ tests with M-As, but also explore a possible new and perhaps superior method of intellectual assessment.

SUMMARY AND CONCLUSIONS

The review of the literature and critique of specific studies presented have shown a great diversity in both approaches and findings concerning the intellectual evaluation of the M-A. Although there tends to be some diversity there are some findings which seem to be emerging with a certain amount of consistency concerning both the state of knowledge in the area and the methodological problems involved in assessing the problem in question.

It appears that hypotheses concerning the intellectual ability of the M-A have been influenced by the research to change. At first studies concerning the M-A used to consider the scores received by the M-A Ss as being truly indicative of their basic intellectual capacity. But now there is a trend to question the IQ tests and to be very careful in the use of their findings. This skepticism of the evaluation of the M-A IQ, by verbal or non-verbal IQ tests standardized on the A-A population, appears to be occurring for two reasons: The first is the observation that verbal and non-verbal tests differ significantly enough in their findings to cast doubt on the verbal tests, and further that even the non-verbal or culture "fair" tests are heavily loaded with items which are actually not so culturally fair. The second reason for skepticism is that teacher judgement or independent measures of ability many times differ radically with IQ test findings.

Hypotheses concerning the M-A have undergone in the past five years a second major change. The changes in the hypotheses have been largely due to the latest research findings which seem to indicate that focusing for representative information, concerning the IQ of M-As on test items, which are not held in common by all groups being studied, will contaminate the results of the tests. The concern then seems to be to find tests or test items which are held in common in the environment

of both the M-A and/or the A-A. Another approach toward overcoming the same problem may be the use of learning tasks standardized on both A-A and M-A populations.

If any summary statement can be made concerning the methodological problems involved in assessing the intelligence of the M-A, it is that the number and variation of factors involved in any study of intellectual ability of the M-A tend to make research control very difficult, and in some cases to date, impossible. There has been an admitted failure by the researchers to control for bilingualism and the process of acculturation in the M-A. Only one approach was demonstrated in the critiques for the control of bilingualism--it was the measure of Language Barrier, described by Holland, (1960). The merits of such a tool for the measure of bilingualism are still to be researched. No specific device or combination of devices has been researched up to this time with the purpose of measuring the amount of acculturation a M-A individual has undergone. This is a specific area which needs to be researched. It would very likely be a most worthwhile contribution to this area if research could be carried out which would clarify many of the problems involved in the intellectual evaluation of the M-A and the acculturation phenomena. The hypotheses which seem to be apparent throughout the research is that M-As differ extensively among themselves in the amount of acculturation they have undergone with respect to the dominant A-A culture. Due to this fact researchers never can be certain of the extent of the differences between the A-A and the M-A groups with which he works.

Controls for rural vs urban, socio-economic status, length of residency in the United States, amount of language barrier, and possibly the acceptance of A-A cultural values, persistently remain as probable contaminating variables operating in the research design. Clarification of the acculturation phenomena may or may

not be included in the exploration of the variables mentioned above, but it does seem that it is closely related to the other variables and may be a part of the same phenomena.

BIBLIOGRAPHY

Altus, G. T., "W.I.S.C. Patterns of a Selective Sample of Bilingual School Children," The Journal of Genetic Psychology, 83, pp. 241-248, 1953.

Anastasi, Anne, Psychological Testing, New York: The MacMillan Company, 1961.

Bingham, C. C., "Intelligence Tests of Immigrant Groups," Psychological Review, 37, pp. 158-165, 1930.

Burns, P. P., The Value of the Leiter Scale in Testing Mexican-American Children, Los Angeles: University of Southern California, Unpublished Master's Thesis, 1954,

Californians of Spanish Surname, Report of the Department of Industrial Relations, San Francisco, 1964.

Carlson, Hilding, B. and Henderson, Norman, "The Intelligence of American Children of Mexican Parentage," Journal of Abnormal and Social Psychology, 45, pp. 544-551, 1950.

Coers, W. C., "Comparative Achievement of White and Mexican Junior High School Pupils," Peabody Journal of Education, 12, pp. 157-162, 1935.

Cohen, J., "A Factor-Analytically Based Rationale for the Weschler Adult Intelligence Scale," Journal of Consulting Psychology, 21, pp. 451-457, 1957.

Cook, J. C. & Arthur, G., 'Intelligence Ratings for 97 Mexican Children in Saint Paul,Minnesota," High School Journal, 38, pp. 24-33, 1955.

Daniel, R. P., "Basic Consideration for Valid Interpretations of Experimental Studies Pertaining to Racial Differences," Journal of Educational Psychology, 23, pp. 15-27, 1932.

Fisher, Gary M., "A Note on the Validity of the Weschler Adult Intelligence Scale for Mental Retardates," Journal of Consulting Psychology, 26, p. 391, 1962.

Goulard, L. J., A Study of the Intelligence of Eleven and Twelve Year Old Mexicans by Means of the Leiter International Performance Scale, Los Angeles: University of Southern California, Unpublished Master's Theses, 1949.

Hill, H. S., "Correlation Between IQs of Bilinguals of Different Ages on Different Intelligence Tests," School and Society, 44, pp. 89-90, 1954.

BIBLIOGRAPHY

(Continued)

Holland, W. R., "Language Barrier as an Educational Problem of Spanish-Speaking Children," Exceptional Children, 27, pp. 42-50, 1960.

Jensen, A. R., "Learning Abilities in Mexican-American and Anglo-American Children," California Journal of Educational Research, 12, pp. 147-149, 1961.

Jones, H. E., "The Environment and Mental Development," In L. Carmichael, Second Edition, Manual of Child Psychology, New York: John Wiley & Sons, pp. 631-696, 1954.

Keston, M. J. and Jimenez, C. A., "A Study of the Performance on English and Spanish Editions of the Stanford-Binet Intelligence Test by Spanish-American Children," Journal of Genetic Psychology, 85, pp. 263-269, 1954.

Kidd, A. H., "The Culture-Fair Aspects of Cattell's Test of g: Culture-Free," Journal of Genetic Psychology, 101, pp. 343-364, 1962.

Koch, H. L. and Simmons, R., "A Study of the Test Performance of American, Mexican and Negro Children," Psychological Monographs, 5, pp. 1-116, 1926.

Mischel, W., "Delay of Gratification, Need for Achievement and Acquiescence in Another Culture," Journal of Abnormal Social Psychology, 62, pp. 543-552, 1961.

Rosen, B. C. & Andade, R. B., "The Psychosocial Origins of Achievement Motivation," Sociometry, 22, pp. 185-218, 1959.

Rosenblum, S., Keller, S. E., & Papania, N., "Davis-Eells (Culture-Fair) Test Performance of Lower-Class Retarded Children," Journal of Consulting Psychology, 19, pp. 51-54, 1955.

Sanchez, G. I., Spanish in the Southwest, Mimeographed, Unpublished Report, 1964.

Silverstein, A. E., Shotwell, A. M., & Fisher, G. M., "Intellectual Functioning of the Mentally Retarded," American Journal of Mental Deficiency, 67, pp. 369-401, 1962

Stablin, J. E., Wiley, D. S. & Thompson, C. W., "An Evaluation of the Davis-Eells (Culture-Fair) Test Using Spanish and Anglo-American Children," Journal of Educational Sociology, 35, pp. 73-78, 1961.

Tireman, L. S., & Watson, M., La Communidad, University of New Mexico Press, Albuquerque, 1943.

Record, W., "Racial Diversity in California Public Schools," Journal of Negro Education, 28, pp. 15-25, 1959.